

Comparing six different species distribution models with several subsets of environmental variables: Predicting the potential current distribution of *Guettarda speciosa* in Indonesia

ANGGA YUDAPUTRA^{1,✉}, INGGIT PUJIASTUTI¹, WENDELL P. CROPPER, JR.²

¹Center of Plant Conservation and Botanic Gardens, Indonesian Institute of Sciences. Jl. Ir. H. Djuanda No.13 Paledang, Central Bogor, Bogor, Indonesia. Tel./fax.: +62-251-8311362-8336871, ✉email: a.yudaputra29@gmail.com

²School of Forest Resources and Conservation, University of Florida, USA

Manuscript received: 10 June 2019. Revision accepted: 26 July 2019.

Abstract. Yudaputra A, Pujiastuti I, Cropper Jr. WP. 2019. Comparing six different species distribution models with several subsets of environmental variables: predicting the potential current distribution of zebra *Guettarda speciosa* in Indonesia. *Biodiversitas* 20: 2321-2328. There are many algorithms of species distribution modeling that widely used to predict the potential distribution pattern of diverse organisms. Finding the best model in terms of predicting the potential distribution of many species remains a challenge. The objective of this study is to compare six different algorithms for predicting the potential current distribution pattern of *Guettarda speciosa* (zebra wood). The occurrence records of *G. speciosa* are derived from herbarium database, Bogor Botanic Gardens's plant inventory database and direct field surveys through NKRI expedition. Seven climatic variables and elevation data are extracted from global data. R open-source software is used to run those algorithms and QGIS is used to prepare the spatial data. The result shows that MAXENT outperforms other predictive models with the highest AUC score 0.89, followed by SVM (0.87), RF (0.86), and GLM (0.82), DOMAIN (0.73), and BIOCLIM (0.62). Based on the AUC score, the four predictive models (MAXENT, SVM, RF, GLM) are categorized into good predictive models, indicating those are quite better to predict the potential current distribution pattern of *G. speciosa*. Whereas, DOMAIN is fair predictive model and BIOCLIM is poor predictive model. The predictive map derived from four models (MAXENT, SVM, RF, and GLM) shows almost similar appearance in predicting of potential current distribution of *G. speciosa*. The predictive map of current distribution would be useful to provide information regarding the potential habitat of *G. speciosa* across the landscape of Indonesia.

Keywords: Algorithms, *Guettarda speciosa*, species distribution modeling

INTRODUCTION

Species distribution models (SDMs) are widely used due to broad applications in ecology, biogeography, and conservation biology. SDMs is also known as habitat suitability models (HSM), Ecological Niche Models (ENM), and Climate Envelope Models. The SDM_s uses multiple environmental predictors for modeling the spatial distribution (presence) of species across landscapes or ecoregions (Segurado and Araújo 2004; Moisen and Frescino 2002). The use of SDM_s has been increasing as they successfully address many problems in ecology such as understanding the ecological requirement of the species (Hirzel et al. 2002), predicting the habitat suitability, assessing biodiversity, managing nature reserves (Araújo and Peterson 2012), predicting the potential for invasion (Zhu et al. 2007) and predicting the current and future distribution pattern of the species under climate change (Gomez and Arriaga 2007).

Many algorithms have been extensively developed to address these issues for better understanding of biodiversity distribution patterns. Those algorithms are classified into regression-based models and machine learning. These methods are able to utilize both presence and absence records and associated background data. BIOCLIM is a classical climate-envelope model that has been widely used

in species distribution modeling (Booth et al. 2014). Even though BIOCLIM does not always perform as strongly as other modeling approaches (Elith et al. 2006), it uses an algorithm that is easily understood and provides a useful insight in methods and procedures of species distribution modeling (Hijmans and Graham 2006). DOMAIN is a simple and robust model for modeling the distribution of plants. The principles of this model are based on a range-standardized and point-to-point similarity metric. An advantage of using DOMAIN is that it operates effectively using presence-only records and relatively few biophysical attributes.

DOMAIN is the heuristic tool that has been identified as suitable for mapping of rare species and for conservation reserve design (Carpenter 1993). In contrast, a recent study found that DOMAIN did not perform very well in a model comparison (Elith et al. 2006). Random Forest (RF) can be applied to both regression and classification problems. The Random Forest technique was designed to provide an accurate prediction, while minimizing overfitting of the data (Breiman 2001a, b). Random Forest modeling applied to species distribution models has been shown to have higher prediction accuracy than a simple decision tree and other models (Prasad et al. 2006; Cutler et al. 2007). Random Forest outperforms compared to other predictive models such as GLM, GAM, MARS, and ANN (Yudaputra

and Robiansyah 2019). Generalized linear model (GLM) has long been used, based on the desire for better and more robust model prediction using independent data, but many algorithms have since been developed that show better performance than GLM (Elith et al. 2006). Support Vector Machine (SVM) uses a simple linear method to reduce the problem of high-dimensional space. The simplicity of SVM combined with many applications (classification, regression, and novelty detection) contributes to the popularity of SVM approach (Karatzoglou et al. 2006). MAXENT (Maximum Entropy) is a machine learning technique that is widely used in species distribution modeling (Phillips et al. 2006). It works very well for the presence-only data (Phillips et al. 2006) and few occurrence records (Pearson et al. 2007). Although any one of these modeling techniques can be applied to species distribution modeling, it is impossible to know which approach will best predict the distribution pattern of any particular species or region. It seems clear that the best approach is to evaluate several different modeling techniques to increase confidence in the ultimate model predictions.

In this study, multiple modeling approaches will be used to predict the potential current distribution pattern of *Guettarda speciosa* (zebra wood) distribution. It is a species of shrub in the family of Rubiaceae, widely distributed across pacific islands, including Micronesia, Fiji, Malaysia, Indonesia and east coast of Africa (GBIF 2018). It reaches the height 6 m, has a fragrant white flower and large green prominently-veined leaves. The common name of the species well known as “jati pasir” in an Indonesian language. However, there are several local names of the species, *laut* (Ambon); *kenyang-kenyang* (Md.); *haruna* (Alor); *tasi* (Timor); *bobalako* (Galela), ketapang pasir (Banyuwangi), Baharu Laut, Kamatolang, Lingka, Pingan-pingan, Salingkaran (Kalimantan) (Heyne 1987; McCormack 2007). It is commonly found in coastal, seashore, sea cliffs, and low land forest habitats in tropical areas. The species has a habitat with the elevation range from 4-50 meter above sea level (McCormack 2007). There

are several utilizations of this species such as the bark is used to treat chronic dysentery and can be applied to wounds, the wood can be used as a raw material for building houses, a dried flower is often used to make essential oil or perfume due to its fragrance (Flora and Fauna Web 2013).

A common challenge to species distribution modeling is how to select the environmental variables used as predictors (Araújo and Guisan 2006). The selection of predictors is fundamentally based on direct and proximal factors that effect on species distribution (Austin 2002), should produce a biologically informative SDM (Newbold 2010). A naïve approach would be to collect as much environmental data as possible and use the model to determine which predictors explain most of the variation. The problems with this approach include resource limitations for data collection, data unavailability, and incomplete ecological knowledge (Bucklin et al. 2015).

Climate is often considered to be the main driver of species distributions. This supported by correlations of past climates with the fossil record (Woodward 1987; Huntley 1999; Davis and Shaw 2000). Examples of the bioclimatic approach include modeling of the distributions of the higher plants of Europe (Beerling et al. 1995; Huntley et al. 1995; Pearson et al. 2002). The hypothesis that continental-scale distributions can be predicted by climate variables, was supported by those studies. Since detailed physiological data for each plant individual is not required, the bioclimatic models can be implemented based on widely available records that include collection locations (Berry et al. 2002). Following the bioclimatic hypothesis, this study uses several climate variables and elevation data to address predicting the distribution of *G. speciosa*.

A robust model of the environmental variables that influence the distribution of *G. speciosa* could have several applications. The model and predicted map could be used by the forest authority and land managers for identifying strategies for conserving the species, e.g., reintroduction and design of optimal conservation areas for protecting the species properly.

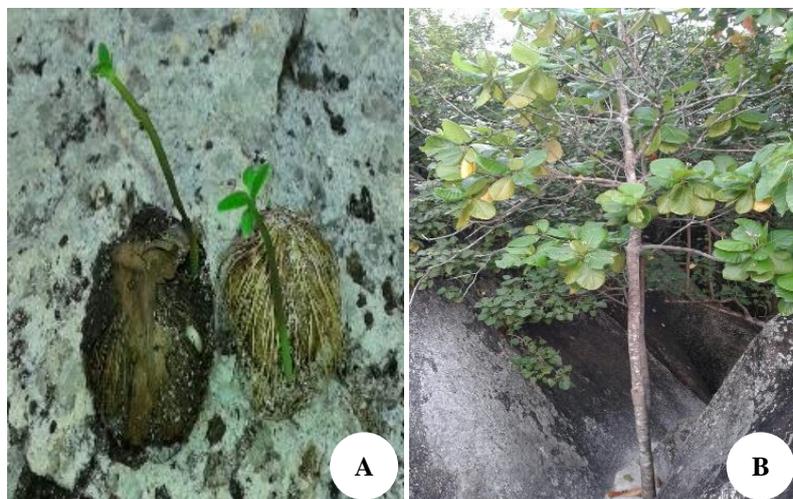


Figure 1. Seedling (A) and mature individuals (B) of *Guettarda speciosa*

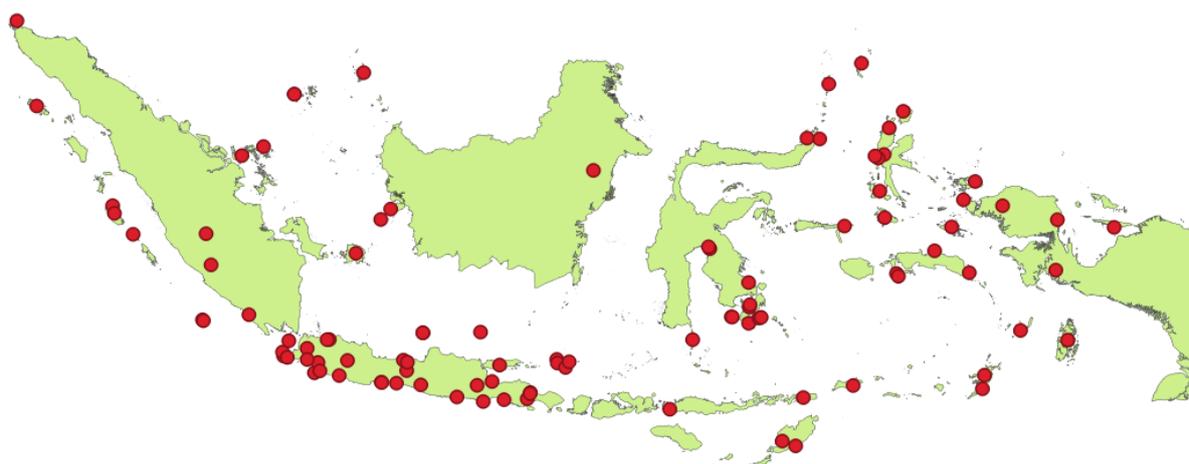


Figure 2. The occurrence records of *Guettarda speciosa* in Indonesia

MATERIALS AND METHODS

The presence records of *Guettarda speciosa* are derived from herbarium collections in the Herbarium Bogoriense, plant inventory database of Bogor Botanic Gardens, and direct observations through the NKRI expedition program in Papua. The expedition is completely carried out at 2016, whereas the herbarium database and the Botanic Gardens's plant inventory database are collected in 2016-2018. Overall, there are 98 occurrence records of *G. speciosa* from all of sources that stated in above (Figure 1).

The climate data is extracted from a global climate data compilation (Worldclim.org). Seven different climate variables are chosen, including annual mean temperature, isothermality, temperature seasonality, annual precipitation, precipitation of the wettest month, precipitation of the driest month, and precipitation seasonality. Elevation is a topographic variable that was used as another input in this model. QGIS is used to organize spatial data.

Six different algorithms are applied to model the pattern of potential current distribution of *G. speciosa*. Those algorithms are BIOCLIM, DOMAIN, RF (Random Forest), GLM (Generalized Linear Model), SVM (Support Vector Machine), and MAXENT (Maximum Entropy). Those models are often used to predict the distribution pattern of species under climate change. All models are run using the R-open source programming language. There are several packages that used in this modeling, include the "bioclim" package used in the BIOCLIM algorithm, the "domain" package for the DOMAIN algorithm, the "randomForest" package for the RF algorithm, the 'glm' package for the GLM algorithm, the 'kernlab' package for the SVM algorithm. The MAXENT model additionally needs java software to run this model. The file "maxent.jar" should be put in the "java" folder of the "dismo" package. This folder is automatically returned by the system.file ("java", package="dismo"). Model performance is then evaluated using the area under the Receiver Operating Curve (AUC).

It is used to evaluate problems regarding to false positives and false negatives in model predictions (Rosset 2004).

RESULTS AND DISCUSSION

Results

Guettarda speciosa is widely distributed throughout Indonesia (Figure 2). The highest frequency of occurrence records is seen in Java, whereas fewer number of occurrence records are seen in Kalimantan. Commonly, this species is found in the peripheral or coastal areas in all regions of mainland Indonesia. Additionally, some occurrence records are found in small islands of Indonesia. The species likely well-growing in coastal areas and lower elevations.

According to model optimization by setting up the k-fold and the number of absence-background data, combination between 4 k-fold and 100 absence-background data showed the highest AUC for all of models. The highest AUC on the best combination of model optimization gained by MAXENT (0.89), followed by SVM (0.87), RF (0.86), GLM (0.82), DOMAIN (0.73), and BIOCLIM (0.62) (Table 1). The AUC for another combination of model optimization is slightly different among each other (Figure 3).

The data is split into test and train data. Train data is used to fix the model performance, whereas the test data is used to evaluate the model. The presence records are split into presence train and presence test. The absence-background data is split into absence-background train and absence-background test.

The spatial current distribution of *G. speciosa* using 6 different predictive distribution models ((Figure 5-10). The map for MAXENT, RF, SVM, and GLM are almost similar, but different in comparing to the predictive map of DOMAIN and BIOCLIM.

Table 1. The AUC score of 6 different species distribution models (BIOCLIM, DOMAIN, GLM, MAXENT, RF, and SVM) in 9 combinations k-fold and the number of absence-background data

Model prediction	AUC (combination k-fold and the number of absence-background data)								
	3 & 1000	3 & 500	3 & 100	4 & 1000	4 k & 500	4 & 100	5 & 1000	5 & 500	5 & 100
DOMAIN	0.64	0.65	0.72	0.67	0.67	0.73	0.65	0.68	0.72
BIOCLIM	0.58	0.59	0.66	0.58	0.57	0.62	0.59	0.61	0.63
GLM	0.75	0.73	0.74	0.80	0.79	0.82	0.79	0.78	0.79
MAXENT	0.82	0.80	0.81	0.86	0.85	0.89	0.86	0.85	0.86
RF	0.80	0.77	0.82	0.81	0.80	0.86	0.84	0.80	0.86
SVM	0.76	0.74	0.82	0.76	0.79	0.87	0.79	0.82	0.85

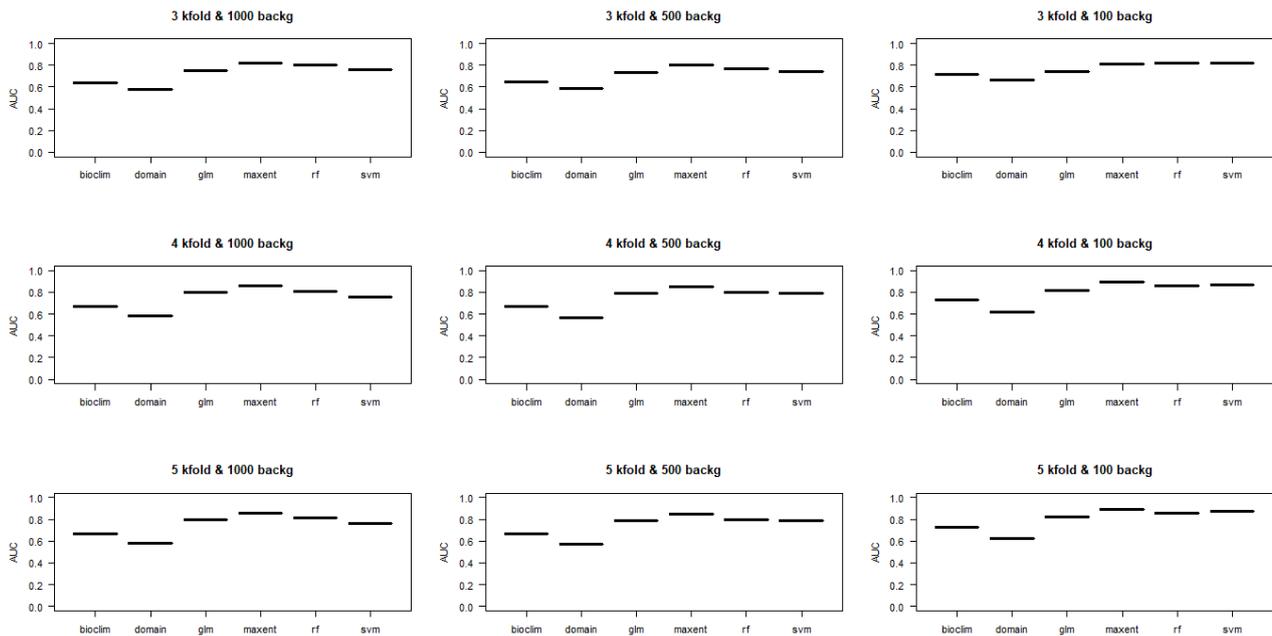


Figure 3. The AUC score of 6 different species distribution models in optimization model performance by setting up the k fold and the number of absence-background data

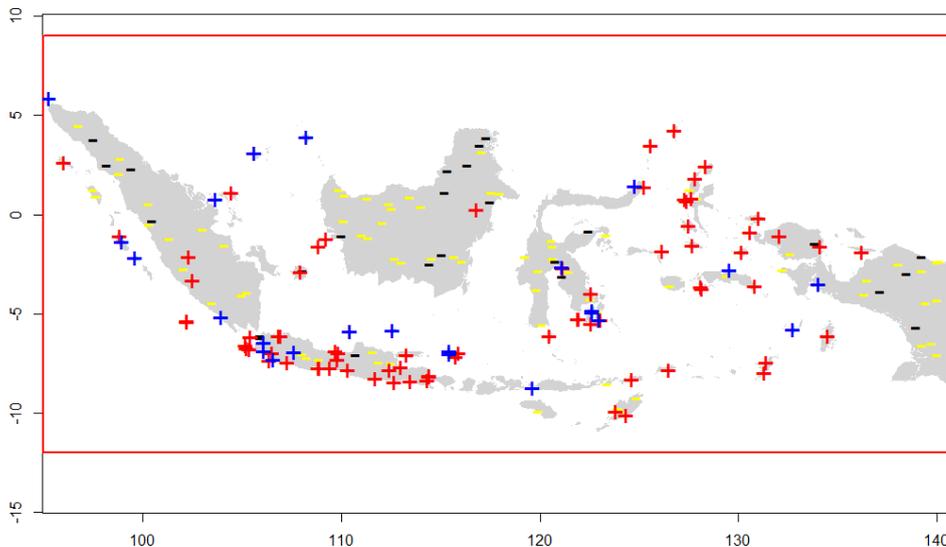


Figure 4. The appearance of model design by splitting data onto training dataset and test dataset, combining presence records and absence-background test. + : presence train, + : presence test,-: absence-background train,-: absence-background test (4 fold and 100 absence-background data)

Discussion

Six different species distribution models (Figures 5-10) are used in this study to predict the current distribution of *G. speciosa*. Those models are BIOCLIM, DOMAIN, GLM, MAXENT, RF, and SVM. Nine combinations of model settings are applied to determine the best model performance. These combinations of settings are constructed by modifying the number of k-fold cross-validations and the number of absence-background data used to find the best model performance represented by AUC. Absence-background data represents the randomly selected of absence occurrence records in geographic area. Those combinations are 3 fold & 1000 absence-background, 3 fold & 500 absence-background data, 3 fold & 100 absence-background data, 4 fold & 1000 absence-background data, 4 fold & 500 absence-background data, 4 fold & 100 absence-background data, 5 fold & 1000 absence-background data, 5 fold & 500 absence-background data, and 5 fold & 100 absence-background data. The model was fit with presence observations for training, presence observations for testing, background testing and background training observations. This model is fit and tested from the splitting data. Calibration and testing data are repeated chosen to produce a more robust model. The splits included presence train: presence test = 75: 25 (4 fold cross-validation). That splitting percentage produced the highest AUC score for all models, even though the AUC score is only slightly different using the 67 :33 (3 fold cross-validation), and 80 : 20 (5 fold cross-validation). In this study, we only used 7 climate predictor variables and 1 topographic predictor variable (elevation). Fielding and Bell (1997) stated the proportion of testing data should be $1/(1 + \sqrt{p} - 1)$, where p is the number of predictors. For example, if there are two predictors, the train : test ratio should be 50 : 50; if there are five, it should be 67 : 33, and if there are many (>10), it should be 75 : 25. Partitioning the data into one training and testing group for model evaluation is an example of two-fold cross-validation. In k-fold cross-validation, the data can be split k times, yielding k estimates of accuracy that can be averaged (Dormann et

al. 2008). In this study, the number of background data affects the AUC score for all of the model predictions. The highest AUC score is found with the model that used 100 background data.

The area under the receiver operating characteristic (ROC) curve (AUC) is used to measure the discriminatory capacity of classification models on species distribution models (SDMs). In the recent decade, the AUC is considered as a standard in measuring the accuracy of species distribution models (Fielding and Bell 1997; Lobo et al. 2008). A model will be considered to discriminate better than chance if the AUC is higher than 0.5 when the curve lies above the diagonal of no discrimination (Krzanowski and Hand 2009). The highest AUC score for six species distribution models is found with the combination of 4 fold & 100 absence-background data. For this combination the resulting AUC scores are as follows: MAXENT has 0.89, SVM has 0.87, RF has 0.86, GLM has 0.82, DOMAIN has 0.76, whereas BIOCLIM has 0.67. This 4 fold & 100 absence-background data combination shows the highest AUC score for all of SDM modeling predictions. In 4 fold & 100 absence-background data combination, MAXENT, SVM, RF, and GLM perform better than other model types for the particular observations used. The highest AUC score is derived from MAXENT, followed by SVM, RF, GLM, DOMAIN, and BIOCLIM. Four species distribution models (MAXENT, SVM, RF, and GLM) have the AUC > 0.80, indicating that those models are good predictive models. DOMAIN belongs to fair model and BIOCLIM to poor model. The value of AUC is calculated by summing the area under ROC Curve, with the range from 0.5 to 1.0. A value of 0.5 is interpreted as no better than a random prediction, a score above 0.5 indicates performance better than random (Hanley and McNeil 1982). The AUC can be used to discriminate between two cases, i.e., species presence and absence (Manel et al. 2001). The score of AUC can represent the category of predictive model, model with AUC in range 0.9-1 (excellent), 0.8-0.9 (good), 0.7-0.8 (fair), 0.6-0.7 (poor), and 0.5-0.6 (fail) (Krzanowski and Hand 2009).

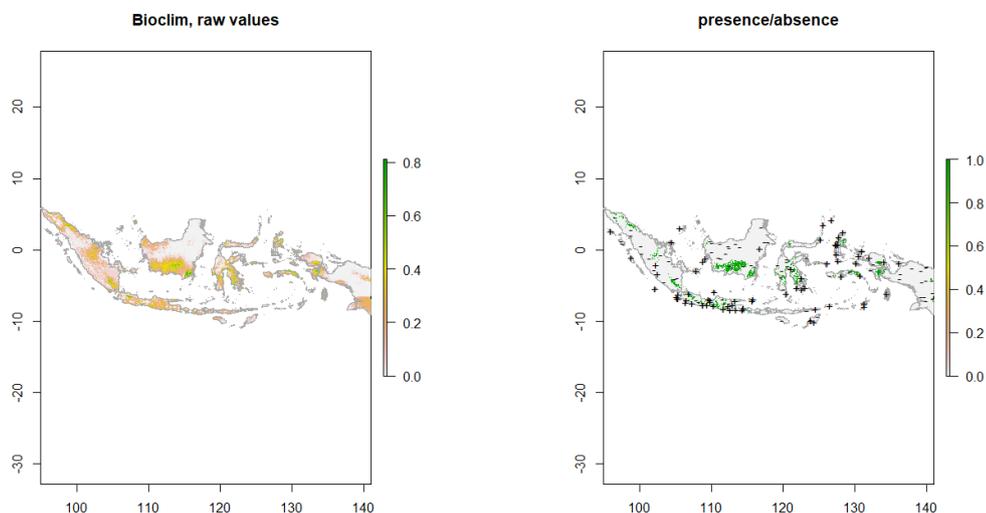


Figure 5. BIOCLIM model prediction (4 fold and 100 absence-background data)

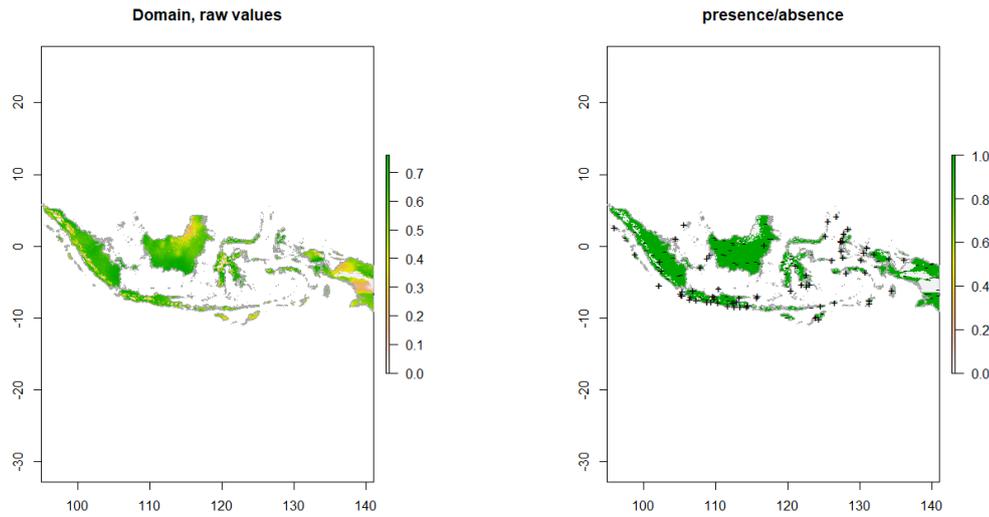


Figure 6. DOMAIN model prediction (4 fold and 100 absence-background data)

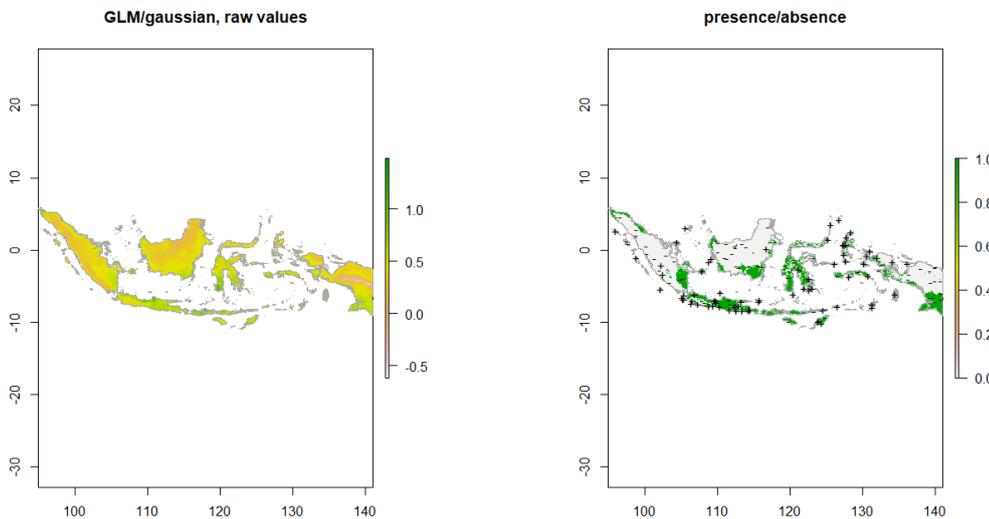


Figure 7. GLM model prediction (4 fold and 100 absence-background data).

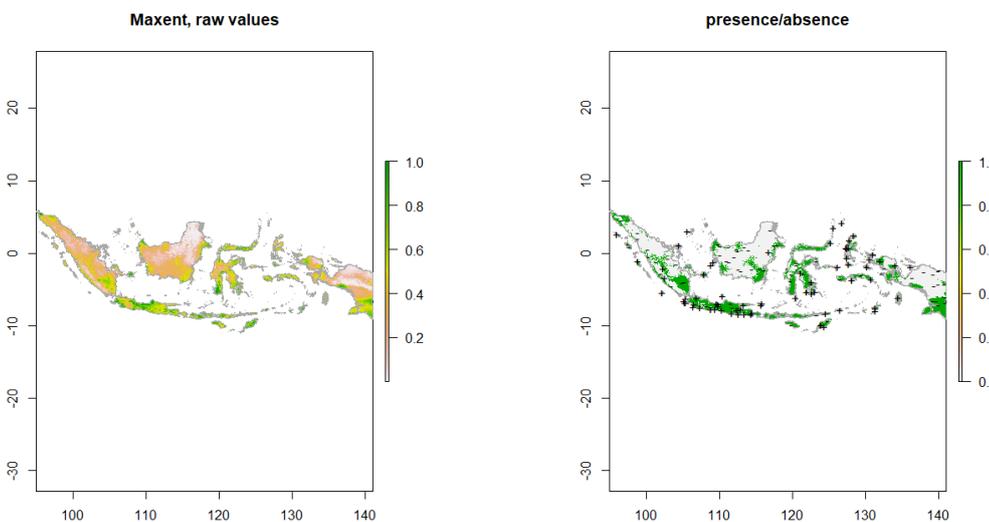


Figure 8. MAXENT model prediction (4 fold and 100 absence-background data).

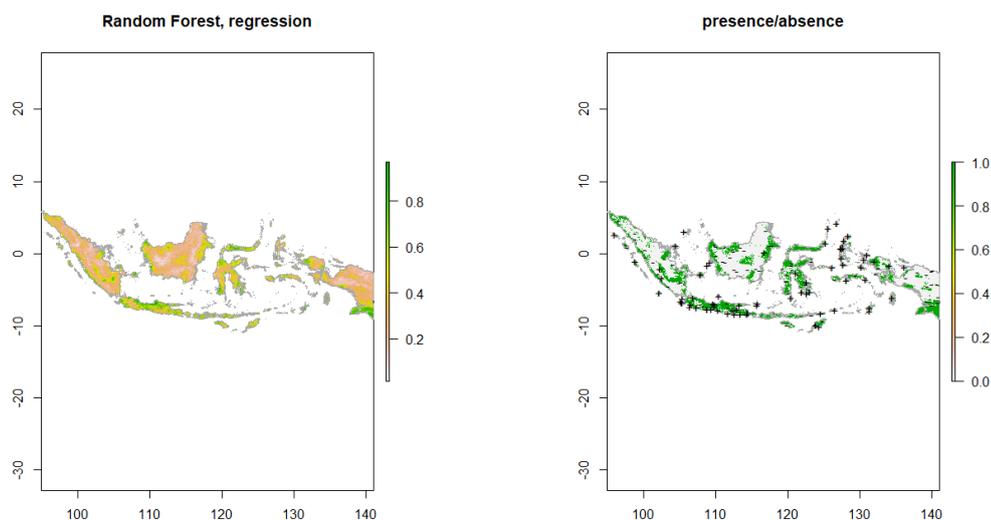


Figure 9. Random Forest (RF) model prediction (4 fold and 100 absence-background data)

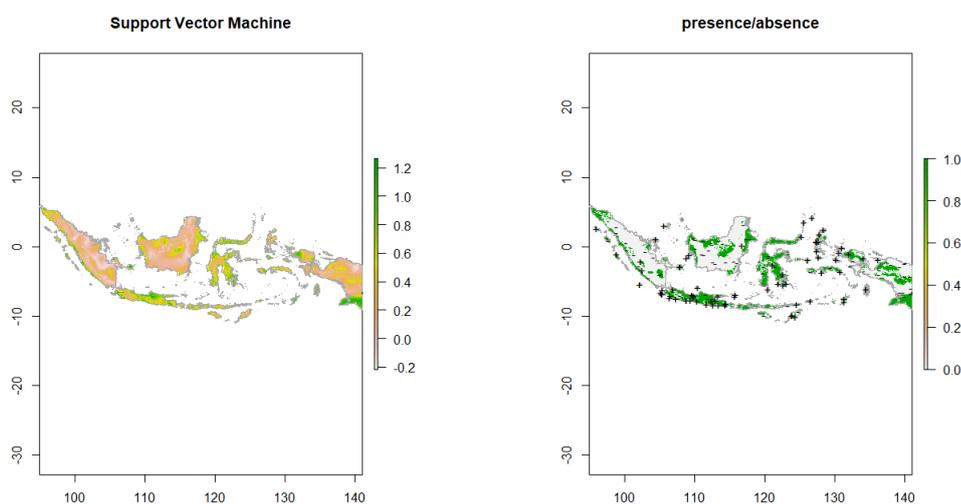


Figure 10. Support Vector Machine (SVM) model prediction (4 fold and 100 absence-background data)

The predictive distribution maps derived from MAXENT, SVM, RF, and GLM models are all almost similar. Those models seem potentially useful for predicting the distribution in coastal regions throughout Indonesia. The predicted maps are quite similar to the observed distribution map (Figures 4-9). Although those models are not categorized as excellent models based on the $AUC < 0.90$, the models have successfully represented the predicted habitat of *G. speciosa*. An $AUC < 0.9$ can occur because of a lack of presence records to cover the huge area with a highly complex environment. This study suggests that models using the MAXENT, SVM, RF, and GLM, can be applied with confidence to map *G. speciosa* habitat in Indonesian coastal areas. The observed $AUC > 80$, is associated with good model performance. These models are largely consistent for locations that *G. speciosa* is predicted to occur and with the observed records of occurrence.

In conclusion, the AUC represents how the performance of predictive models in terms of predicting the potential current distribution of *Guettarda speciosa*. The higher AUC score, the better model will be performing. MAXENT shows the highest AUC score, followed by SVM, RF, GLM, DOMAIN, and BIOCLIM. Those four models (MAXENT, SVM, RF, and GLM) are categorized into good predictive models. Whereas, DOMAIN is fair model and BIOCLIM is poor model in case of predicting potential current distribution of *G. speciosa*.

ACKNOWLEDGEMENTS

We would like to express a highly appreciated and thankful to staff of Herbarium Bogoriense and Bogor Botanic Gardens, Indonesia in terms of providing useful information about the herbarium specimen and database of *Guettarda speciosa*.

REFERENCES

- Araújo MB, Guisan A. 2006. Five (or so) challenges for species distribution modelling. *J Biogeogr* 33: 1677-1688.
- Araújo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* 93: 1527-1539.
- Austin M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol Model* 157: 101-118.
- Beerling DJ, Huntley B, Bailey JP. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *J Veg Sci* 6: 269-282.
- Berry PM, Dawson TP, Harrison PA, Pearson, RG. 2002. Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Glob Ecol Biogeogr* 11: 453-462.
- Booth TH, Nix HA, Busby JR, Hutchinson MF. 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Divers Distrib* 20: 1-9.
- Breiman L. 2001a. Statistical Modeling: The Two Cultures. *Statistical Sci* 16: 199-215.
- Breiman L. 2001b. Random Forests. *Machine Learnin* 45: 5-32.
- Bucklin DN, Basille M, Bencotter AM, Brandt LA, Mazzotti FJ, Romanach SS, Sperotera C, Watling JI. 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Divers Distribu* 21 (1): 23-35.
- Carpenter G, Gillison AN, Winter J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers Conserv* 2: 667-680.
- Cutler DR, Edwards, TC, Beard KH, Cutler A, Hess KT, Gibson JC, Lawler JJ. 2007. Random forests for classification in ecology. *Ecology* 88(11): 2783-2792.
- Davis MB, Shaw, RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673-679.
- Dormann CF, Fründ J, Gruber B. 2008. Introducing the bipartite Package: Analysing Ecological Networks. <https://www.researchgate.net/publication/228861770>.
- Eliith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick J, Lehmann A, Li J, Lohmann LG, Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. Overton J, Peterson AT, Phillips S, Richardson K, Scachetti-Pereira R, Schapire R, Soberon J, Williams S, Wisz M, Zimmerman N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Fielding AH, Bell JF. 1997. A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models. *Environ Conserv* 24: 38-49.
- Flora and Fauna Web. 2013. National Parks Board. Singapore. <https://florafaunaweb.nparks.gov.sg/Special-Pages/plant-detail.aspx?id=2946>.
- Gómez-Mendoza L, Arriaga L. 2007. Modeling the effect of climate change on the distribution of oak and pine species of Mexico. *Conserv Biol* 21: 1545-1555.
- GBIF. 2018. *Guettarda speciosa*. GBIF Occurrence (GBIF.org). <https://doi.org/10.15468/dl.ncywn>.
- Hanley JA, McNeil BJ. 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143: 29-36.
- Heyne K. 1987. *Tumbuhan Berguna Indonesia III*. Badan Litbang Kehutanan, Departemen Kehutanan, Jakarta. [Indonesian]
- Hijmans RJ, Graham CH. 2006. Testing the ability of climate envelope models to predict the effect of climate change on species distributions. *Glob Change Biol* 12: 2272-2281.
- Hirzel A, Hausser J, Chessel D, Perrin N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-2036
- Huntley B, Berry PM, Cramer W, McDonald AP. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *J Biogeogr* 22: 967-1001.
- Huntley B. 1999. Species distribution and environmental change: considerations from the site to the landscape scale. In: Maltby E, Holdgate M, Acreman M, Weir A (eds) *Ecosystem management: questions for science and society*. Royal Holloway Institute for Environmental Research, Virginia Water, UK.
- Karatzoglou A, Meyer D, Hornik K. 2006. Support Vector Machines in R. *J Statistical Software* 15(9). <http://www.jstatsoft.org/v15/i09/>.
- Krzanowski WJ, Hand DJ. 2009. ROC curves for continuous data. Chapman and Hall, Boca Raton, FL.
- Lobo JM, Jiménez-Valverde A, Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17: 145-151.
- Manel S, Ormerod SJ, Williams HC. 2002. Evaluating presence-absence models in ecology: The need to account for prevalence. *J Appl Ecol* 38(5): 921-931.
- McCormack G. 2007. *Guettarda speciosa*. Bishop Museum, Cook Island Biodiversity Database.
- Moisen GG, Frescino TS. 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecol Model* 157: 209-225.
- Newbold T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geogr* 34:3.
- Pearson RG, Dawson TP, Berry PM, Harrison PA. 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecol Model* 154: 289-300.
- Pearson RG, Nakamura M, Raxworthy CJ, Peterson AT. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J Biogeogr* 34(1):102-117.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model* 190: 231-259.
- Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression techniques: Bagging and random forests for ecological prediction. *Ecosystem* 9: 181-199.
- Rosset S. 2004. Model selection via the AUC. In *Proceedings of the twenty-first international conference on Machine learning*: 89.
- Segurado P, Araujo MB. 2004. An evaluation of methods for modelling species distributions. *Biogeogr J* 31: 1555-1568.
- Woodward FI. 1987. *Climate and plant distribution*. Cambridge University Press, Cambridge.
- Yudaputra A, Robiansyah I. 2019. The Implementation of Artificial Neural Network and Random Forest in Ecological Research: Species Distribution Modelling with Presence and Absence Dataset. *Proceedings SATREPS Conference*.
- Zhu L, Sun OJ, Sang W, Li Z, Ma K. 2007. Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. *Landsc Ecol* 22: 1143-1154.